

Imagining Experiences

Peter Langland-Hassan

University of Cincinnati

Abstract: It is often held that in imagining experiences we exploit a special imagistic way of representing mentality—one that enables us to think about mental states in terms of *what it is like* to have them. According to some, when this way of thinking about the mind is paired with more objective means, an explanatory gap between the phenomenal and physical features of mental states arises. This paper advances a view along those lines, but with a twist. What many take for a special imagistic way of thinking about experiences is instead a special way of misconstruing them. It is this tendency to misrepresent experiences through the use of imagery that gives rise to the appearance of an explanatory gap. The pervasiveness and tenacity of this misrepresentational reflex can be traced to its roots in a particular heuristic for monitoring and remembering the mental states of others.

I. Imagining the gap

There are limits to what we can imagine.

A case in point: Thomas Nagel's cognitive resources are "inadequate to the task" of imagining a bat's experiences. "Nothing in my present condition," he says, "enables me to imagine what the experiences...would be like" (1974, p. 439). If we *could* imagine a bat's experiences, Nagel suggests, we would understand things about the mental lives of bats that we do not currently understand. But, intuitively, we cannot imagine a bat's experiences.

Similarly, Frank Jackson's (1982) Mary does not know what it is like to see red before leaving her black and white room. If she could imagine the experience of seeing red, she would have the knowledge in question. But, intuitively, she cannot.

These imaginative blocks trace, at least in part, to an inability to form the right kind of sensory imagery. Mary cannot imagine the experience of seeing red because she cannot form the necessary visual image. And we cannot imagine a bat's experiences of moths, mosquitos, and the like because we cannot generate the right kind of echolocative imagery. Forming the

proper sort of mental image may not *suffice* for engaging in the imaginings in question. But, I take it, most will agree it is at least necessary.

Thus it is the *imagistic* or *sensory* imagining of experiences that seems to be of a special sort. To many, it seems to offer a distinctive way of knowing about experiences. A constellation of terms has been used to describe what is special about it. It has been said that imagery allows us to imagine experiences “from the inside” (Levin, 2007a; Noordhof, 2002; Peacocke, 1985), or to think about mentality in a *first-personal* way (Paul, 2014), grasping the *subjective* (Nagel, 1974) or *phenomenal* (Balog, 2012; Block, 2006; Papineau, 2002, 2007b) features of experience. This sort of understanding is sometimes contrasted with more objective means for thinking about minds, such as those deriving from the cognitive sciences.

Many who assign special significance to the ability to imagine experiences have also thought that the way we use mental imagery when considering the mind/brain relation has something to do with the appearance of an *explanatory gap* between brain processes and consciousness (Balog, 2012; Block, 2006; Hill, 1997; Levin, 2007a; Nagel, 1974; Papineau, 2002, 2007b). Nagel was one of the first to develop a proposal of this kind in a well-known footnote to his 1974, where he speculates that even if mental states were identical to brain states, the two would seem separable due to the different ways in which we imagine mental and brain states (1974, fn. 11). In a similar vein, some have held that mental images, and other states with phenomenal character, form proper parts of the concepts we use to think about states with phenomenal character. It is the peculiar fact that such *phenomenal concepts* contain instances of their referents that, for the views in question, accounts for the appearance of an explanatory gap. These have become known as *constitutional accounts* of phenomenal concepts (Balog, 2012; Block, 2006; Hill & McLaughlin, 1999; Papineau, 2002, 2007b).

Both in Nagel’s brief footnote and the more recent constitutional accounts of phenomenal concepts, there are two separate claims worth distinguishing. The first is that the use of mental imagery in thoughts about experiences gives rise to the appearance of¹ an explanatory gap between conscious mental states and brain processes. The second is that, when we use imagery to think about experiences, our thoughts are generally *veridical*, insofar as we do not tend to misrepresent the experiences. This *veridicality assumption* is worth

¹What matters for present purposes is simply the claim that imagery is involved in generating the *appearance* of an explanatory gap of the kind described by Levine (1983), whether or not one thinks there will always be such a gap in our understanding, and whether or not one thinks the appearance of such a gap is veridical.

highlighting. For one *might* have thought that, if there is an unusual gap in our understanding created through the disparity of two means of thinking about the mind, it is because one of the ways is getting things wrong—just as apparently conflicting reports about some event might lead us to think that one of the reports is false. But this is not how the idea is typically developed. Instead it is suggested that, while sympathetic imagination (in Nagel’s case) and phenomenal concepts (for defenders of constitutional accounts) provide us with a substantive and veridical picture of mentality, the way they acquaint us with the mind is so different from other modes of thought that it seems we are thinking of two different entities, contingently related, when we are in fact thinking of one and the same thing in two different ways (Nagel, 1974, fn. 111; Hill & McLaughlin, 1999; Balog, 2012). It is the failure of (what we might loosely call) *conceptual alignment* between two otherwise veridical conceptions of one and the same thing that, for the views in question, accounts for the explanatory gap.

This paper proposes a different sort of link between mental imagery and the explanatory gap—one which challenges the veridicality assumption. My argument will be that mental imagery is often used in a way that *misrepresents* the nature of experiences, and that it is this common *misuse* of imagery that partly accounts for the appearance of an explanatory gap. (I say “partly accounts” because I doubt there is just *one* reason a person might intuit a mind-brain gap.) In developing this argument, I want to put pressure on the very idea that there is a special imagistic way of thinking about the subjective, phenomenal features of mental states.

My strategy will be first to develop (in Section 2) a general challenge to the idea that there is a special imagistic way of thinking about experiences, by showing that existing accounts have not adequately specified the difference between using imagery to think about experiences in terms of what it is like to have them, and using imagery to think about experiences in terms of their perceptible features. This challenge is specifically aimed at Nagel’s distinction between sympathetic and perceptual imagination; yet, if cogent, it extends to existing constitutional accounts of phenomenal concepts (e.g., Balog (2012), Block (2006), Papineau (2007b)), and, indeed, to any theory which posits a special imagistic way of thinking about experiences.²

² The argument does not, however, take aim at demonstrative accounts of phenomenal concepts, such as those of Loar (1997) and Levin (2007b). Demonstrative accounts highlight the use of a demonstrative in phenomenal concepts, and not the use of a sensory image. Thus, they do not posit a special *imagistic* way of thinking about experiences but, rather, a special *demonstrative* way of thinking about experiences. My target here is the idea that mental images are not merely the *referents* of some phenomenal concepts, but are also *proper parts* of the referring concepts or thoughts themselves that we use when we think about mental states.

While the argument falls short of a *demonstration* that there is no special imagistic way of thinking about experiences in terms of what it is like to have them, it opens the door to better-specified alternatives. The subsequent section (Section 3) then defends an alternative proposal for what is involved in imagining experiences. This account is able to answer the challenge raised in the previous section, concerning the difference between using imagery to think about experiences in terms of what it is like to have them, and using imagery to think about experiences in terms of their perceptible features. It is from within the terms of this approach that an account of the explanatory gap is available. I argue that the gap arises, in part, from a tendency to misattribute to experiences properties of a kind they lack. The pervasiveness and tenacity of this misrepresentational reflex is then (in Section 4) traced to its roots in a particular heuristic for monitoring and remembering the mental states of others.

II. Two ways of imagining?

Thomas Nagel's distinction between perceptual and sympathetic imagination is a good place to begin in considering what it is to imagine experiences, both because it is influential and because the problems it faces extend to more contemporary views that take inspiration from his account (or so I will argue).

In a now well-known footnote to his 1974, Nagel proposes that the reason mental states and brain states would seem only contingently related, even if they were identical, is that they are imagined in two very different ways: perceptually and sympathetically. To imagine something perceptually, Nagel writes, "we put ourselves in a conscious state resembling the state we would be in if we perceived it." In perceptually imagining a toaster, for instance, we might form a visual image of a toaster. For that is to put ourselves in a state resembling the state we would be in if we perceived a toaster. By contrast, to imagine something sympathetically, "we put ourselves in a conscious state resembling *the thing itself*" (p. 466, fn. 11, emphasis added). An important caveat concerning sympathetic imagination is that it "can be used only to imagine mental events and states—our own or another's" (ibid.). Thus, there can be no sympathetic imagining of a toaster, because we cannot enter into a mental state that resembles a toaster itself—at least, not one that resembles a toaster in the right way.³ We can,

³ Nagel is not very explicit concerning the *respects in which* states must resemble each other for each kind of imagination to occur, specifying only, for sympathetic imagination, that the states resemble each other "mentally". But it seems safe to assume that for both kinds of imagination the resemblances he has in mind are with respect to

however, sympathetically imagine a *visual perceptual experience* of a toaster, as the experience of a toaster is a kind of mental state. To do so, we put ourselves in a state resembling the visual experience itself. This might also involve forming an image of a toaster, since to form such an image is to enter a mental state resembling “the thing itself,” namely, the visual experience of a toaster.

Applying this distinction to the mind-body problem, Nagel proposes that we typically imagine mental properties in one way (sympathetically), and the physical properties with which they may be identical in another way (perceptually). (Note: Nagel uses the term ‘features’ in place of ‘properties’). “Where the imagination of physical features is perceptual and the imagination of mental features is sympathetic,” he explains, “it appears to us that we can imagine any experience occurring without its associated brain state, and vice versa. The relation between them will appear contingent even if it is necessary, because of the independence of the disparate types of imagination” (Nagel, 1974: 446, fn. 11).

Suppose, for instance, that some version of the identity theory is true, such that brain state *B* is token identical with visual experience *E*. Visual experience *E*, we can suppose, is a visual experience of a beach ball. While *B* is identical to *E*, we are nevertheless inclined to imagine them in very different ways (according to Nagel). In sympathetically imagining *E*, we form a visual image of a beach ball. For in entering this mental state we enter a state that resembles *E* itself. On the other hand, we are inclined to imagine brain state *B* perceptually, as brain states are ordinarily conceived as things in the world which, under the right circumstances, we might perceive. To perceptually imagine *B*, we *do not* form a visual image of a beach ball. For *that* visual image does not resemble that kind of state we would be in if we perceived a brain state. Instead, we form a visual image of a certain kind of (living, functioning) brain state. Such an image presumably resembles the kind of state we would be in if we in fact perceived a certain brain state.⁴

intentional and/or phenomenal properties, as these are the most salient properties of mental states. (We can remain neutral here on the relationship between intentional and phenomenal properties).

⁴ In an influential development of Nagel’s idea, Christopher Hill holds that brain processes are not likely the sort of things that are directly perceivable—that they probably fall “on the theoretical side of the fuzzy line that divides theoretical entities from observable entities (1997, p. 68). I disagree. I think that brain states can in fact be *seen*, in the right circumstances. Brain states are anatomical states after all, not microphysical ones. Neurosurgeons are able to observe them. And as is the case with natural kinds in general, one need not be able to perceptually discriminate the features that make a brain state the kind of state it is, in order to perceive it. Nor must one perceive the *entire* brain state to count as perceiving it. That said, if one agrees with Hill that brain states are not

Now, to the extent that the beach ball image used in sympathetically imagining *E* and the brain state image used in perceptually imagining *B* seem arbitrarily related—and why wouldn't they?—it will seem odd to suppose that we are in fact imagining one and the same thing with each image. It is not as though the images seem to be of the same thing but from different angles, or from up-close and far away. Nor is either of the images thought to achieve reference through mere stipulation or metaphor. Thus, even if we lean toward accepting an identity between the two objects of imagination, they continue to appear entirely distinct, with no logical path from one to the other. This makes the putative identity intuitively puzzling. Indeed, there is no *greater* intuitive discomfort in sympathetically imagining *E* while perceptually imagining the complete absence of *B*, or vice versa (Cf. Hill, 1997, p. 69-70). This, I take it, is how Nagel's proposal is standardly understood.

Nagel's diagnosis of contingency intuitions has been criticized as too general, insofar as the imaginative mechanisms he appeals to ought to generate intuitions of contingency we lack (Doggett & Stoljar, 2010; Papineau, 2007a; Yablo, 2006). And it has been found too narrow in that there are mistaken intuitions of contingency it does not adequately explain (Hill, 1997; Doggett & Stojar, 2010). Yet, to my mind, these objections deal only a passing blow. We should not deem a strategy for explaining particular mistaken modal intuitions a success only if it can be generalized to explain every mistaken modal intuition. The different forms of imagination might contribute to a sense of contingency concerning the *mind-brain* relation, even if there are other mistaken intuitions of contingency it cannot explain. Nor should we assume that Nagel's diagnosis is undermined if there are cases where we lack an intuition of contingency, but where image-splicing of the kind he describes could conceivably be used to generate such an intuition. For there may be reasons we are more inclined to take seriously the intuitive results of our image-splicing in the mind-brain case than in other contexts (one such reason is explored in Section 4).

My aim, however, is not to defend Nagel's diagnosis, but to offer a deeper criticism. For even abandoning Nagel's proposal *qua* explanation of mistaken intuitions, one might still think that sympathetic imagination offers a special form of *insight* into the nature of mentality—one that reflects our knowledge of aspects of consciousness that (perhaps) do not properly mesh

in fact perceivable, one can follow Hill in holding that perceptually imagining *B* involves forming an image of an apparatus or model that allows one to indirectly perceive *B* (Cf. Hill (1997, p. 68-69)).

with the mind as understood by the natural sciences. Nagel, for one, thinks that we will not truly *understand* how mind-brain identities are possible without a Copernican revolution in our conceptual scheme (1974, p. 447-8). The felt need for this revolution can be traced to the sort of knowledge that sympathetic imagining is thought to provide. After all, if what there is to be known through sympathetically imagining a bat's experiences could be deduced from studying its neuroanatomy, the revolution could be called off. It is the very existence of this special means for thinking about mentality that I want to question.

2.1 *Instability in the Sympathetic/Perceptual distinction*

Despite its intuitive appeal, Nagel's distinction between sympathetic and perceptual imagining begins to dissolve under scrutiny. Suppose, returning to our earlier example, that we want to perceptually imagine an actual beach ball, and not a brain state. This can be done by putting ourselves in a conscious state that resembles brain state *B*, because *B* is the kind of state one would be in if one in fact perceived a beach ball. Yet we already saw that *sympathetically* imagining *B* is *also* accomplished by entering into a mental state that resembles "*B* itself" (on the supposition that $B=E$). This entails that perceptually imagining a beach ball is *also* sympathetically imagining brain state *B*. And, of course, when we sympathetically imagined *B* (by entering into a mental state that resembled *B* itself), we were also entering into a mental state that resembles the state we would be in if we were to perceive a beach ball. Thus we were also (unwittingly) perceptually imagining a beach ball.

Indeed, taking Nagel's way of explaining the distinction at face value, sympathetically imagining *any* mental state that represents objects or properties of the kind we could perceive will also, simultaneously, be a case of perceptually imagining those objects or properties. And, by the same token, perceptually imagining any object will be a case of sympathetically imagining the mental state one would be in if one perceived that object. If all cases of perceptually imagining are indeed simultaneously cases of sympathetic imagining, then it is wrong to conceive of them as two different *kinds* of imagination, where an instance of one type combines with an instance of the other to generate an appearance of contingency. Nor can sympathetic imagining be counted as offering a special way of thinking about the phenomenal features of experiences—a way not available through perceptual imagination—if most or all instances of one form are also instances of the other. (Whether the answer is "most" or "all" here

depends on one's view of bodily sensations.⁵) And even if *some* sympathetic imaginings are not also perceptual imaginings (a view which I reject in fn. 5), the fact that all perceptual imaginings are sympathetic imaginings is sufficient by itself to show that there are not two importantly distinct classes of imaginings here, involving different uses of imagery; one remains a subset of the other.⁶

To be clear, the problem is not that two different imaginative projects are shown to make use of the same type of mental image. One could grant that much but insist that, during instances of the two kinds of imagining, the same *type* of image may be put to use in imagining different *objects*. The problem is rather that, *given Nagel's way of defining each kind of imagination*, instances of perceptual imagining *just are* instances of sympathetic imagining. We do not have two separate groups of mental acts, some of which are sympathetic imaginings while the others are perceptual. Rather, we have single, token acts that are *both* sympathetic imaginings and perceptual imaginings. This is because any token mental state that resembles the state you would be in if you perceived, e.g., a beach ball will also resemble the perceptual experience as of a beach ball. In fulfilling the criterion for being a perceptual imagining, the same token state simultaneously fulfills the criterion for being a sympathetic imagining (as Nagel defines these

⁵ Whether all cases of sympathetic imaging are also cases of perceptual imagining depends on whether there are sensory states that are not involved in perception; if there are, then cases of sympathetically imagining those states will not be cases of perceptual imagining. One might think that pains are examples of such states. However, many (including myself) view pain as a form of bodily perception, or "interoception" (Armstrong, 1962; Bain, 2007; O'Sullivan & Schroer, 2012; Tye, 2005). It is because of pain's perceptual character that we can say, for instance, that a person suffering phantom limb pains is *misperceiving* his limb as being a certain way. With this view of pain in place, sympathetically imagining a pain becomes a case of perceptually imagining a state of one's body (such as a state of tissue damage or nerve stimulation). On the other hand, just to be clear, perceptually imagining a pain—say, by entering into a state similar to the one you would be in if you were to see a certain neural state identical to a pain—is *not* the same thing as sympathetically imagining a pain. Yet, it *is* the same things as sympathetically imagining the visual experience as of a pain. (Pains are potentially visible for the same reason that any mental state that is identical to some brain state is potentially visible—see fn. 4).

⁶ Stoljar (2006) (pp. 191-192) seems to have the same sort of criticism in mind in arguing that Nagel and Hill's distinction between two kinds of imagination collapses. Yet his version of the criticism relies on a subtle mischaracterization of their notion of perceptual imagination. Stoljar glosses perceptual imagination as requiring one "to imagine oneself perceiving or apparently perceiving that such and such is the case." This, he says, is simply "to imagine oneself being in a conscious state," which, he observes, is the same as sympathetically imagining that state. Together this entails that "imagining oneself perceiving something [i.e. perceptual imagination] is sympathetic imagination" (p. 192). Yet neither Nagel nor Hill holds that in perceptually imaging an object one "imagines oneself perceiving, or apparently perceiving" the object. They simply hold that you enter into a conscious state similar to one you would be in if you perceived such an object or property. This is not at all equivalent to imagining *yourself perceiving* the object. The latter involves thinking about yourself as being in a certain mental state, the former does not. Without this subtle mischaracterization of perceptual imagination, Stoljar's argument does not go through.

notions). And if every token perceptual imagining is already a sympathetic imagining, it cannot be that images are put to use in one way during perceptual imaginings, and another way during sympathetic imaginings. Nor can it be that one kind of imagining has one sort of object (an experience) while the other has another (a worldly object one might perceive). For each token perceptual imagining, being also a sympathetic imagining, is used in the same way as a sympathetic imagining; so there are not two different uses that could warrant speaking of two different objects.

While faithful to Nagel's text, this criticism may nevertheless appear uncharitable. Nagel's footnote is but a footnote, after all. Perhaps we should not expect it to offer an airtight account of the intuitive distinction it introduces. Yet, charitable or not, the criticism applies equally to Christopher Hill's (1997) more carefully developed account of Nagel's perceptual/sympathetic distinction, as it also grounds the distinction in resemblance relations.⁷ That said, the important question going forward is whether there is indeed some better way of articulating the intuitive distinction that Nagel (and Hill) are after that preserves its *core idea*. And I take it the core idea is that sympathetic and perceptual imagination offer conceptually or cognitively distinct (yet veridical) ways of thinking about mental states, where the distinctness of imaginative modes is apt to generate the appearance of an explanatory gap. How else might we understand these different ways? That is the question to which I turn now.

2.2 Cognitively equivalent perceptual and sympathetic imaginings

We already know that if the same type of image can be used in both a sympathetic and perceptual imagining, then the question of which type of imagining a person is undergoing cannot be settled by appeal to the type of image in use. In virtue of what, then, does an imagining count as either a case of perceptual or sympathetic imagination? An attractive answer that fills the apparent gap in Nagel and Hill's account of the distinction might appeal to the *purposes or uses* to which the image is being put by the imager, by building those purposes or

⁷ "If P is a property of which one can be introspectively or perceptually aware," writes Hill, "then, when one imagines an instance of P, what one does is to put oneself into a state which is similar to the state one is in when one is experientially aware of an instance of P" (1997, p. 66, emphasis added). Presumably, the state one is in when one is "experientially aware" of the property of being a cube is similar to the state one is in when one is experientially aware of the phenomenal properties of a visual experience of a cube. Thus, an imagining that is similar to one will be similar to the other, making the imagining an instance of both kinds of imagination.

uses into the very definition of the kinds of imagination. (Recall that Nagel and Hill only appeal to brute resemblances in their criteria; the imaginers' purposes do not enter into the criteria). Roughly, we might say that perceptually imagining an *X* occurs when, *for the purpose of thinking about some X we could perceive*, we form a mental image that resembles the state we would be in if we perceived an *X*. Sympathetically imagining *Y*, by contrast, might occur when we form a mental image that resembles mental state *Y*, *for the purpose of thinking about Y itself*. For even if a certain mental image resembles both mental state *Y* itself and the kind of state one would be in if one perceived an *X*, there might be *functional differences* in how the image is used within each kind of imagining, deriving from the different *purposes* characteristic of each. These functional differences could conceivably ground the distinction between two different kinds of imagining and give life to the idea that the two kinds of imagining have different objects when they nevertheless make use of the same type of image.

Yet this amendment does not in fact preserve the core idea behind the distinction, if we suppose (with the identity theorist) that mental states are brain states and, therefore, parts of physical reality that can, under the right conditions, be perceived. To see why, suppose that someone (call him Owen) is undergoing an awake craniotomy. An awake craniotomy is a form of brain surgery where the patient remains conscious throughout the operation. By administering only a local anesthetic, surgeons are able to monitor the patient's cognitive capacities throughout the operation. Understandably, patients undergoing an awake craniotomy do not, as the operation is ongoing, look at the portions of their brain that are being operated upon.⁸ But there is no reason the brave and curious could not do so through an arrangement of mirrors. With that in mind, consider a hypothetical case:

Owen's Operation: Owen is watching his own awake craniotomy. He is looking, through mirrors, at the parts of his brain responsible for his current visual experience *E*, as a nearby tumor is resected. *E* represents brain state *B*, which is giving rise to *E* itself and which (let us suppose) is token identical with *E*. Now Owen decides to sympathetically imagine *E*. To do so, he forms visual image *M*, which resembles *E*, for the purpose of thinking about *E* itself. Following that, he decides to perceptually imagine *B*.

⁸ Though some have come close. See, e.g., Charles Trippy's video of himself watching his own awake craniotomy after the fact: <https://www.youtube.com/watch?v=fCAAe0ROMp4>.

To do so, he forms a visual image that resembles the state he would be in if he perceived *B*, for the purpose of thinking about *B* (in line with our modified definition above). Of course, *E* just is the kind of state he would be in if he perceived *B*. So again he forms visual image *M*, which resembles *E*.

Given our assumption that $E=B$, we have a case where an imagining of each kind—sympathetic and perceptual—makes use of the same type of image in an imagining of *the same thing* (namely, E/B). We can add, for good measure, that Owen believes that *B* is identical to *E*. We are, after all, attempting to account for an intuition of contingency that seems to linger *even after* we accept mind-brain identities. In that case, his purposes converge as well: for him to form an image of type *M* for the purpose of thinking about some *B* he could perceive is cognitively equivalent to forming an image of type *M* for the purpose of thinking about *E* itself. For he conceives of *B* and *E* as being the same. Thus, in Owen's imaginings, we have a convergence of three things: that which is imagined (i.e. B/E), the type of image that is used, and the purpose for which the image is used.

Can we nevertheless say that Owen has two different kinds of thoughts here—two different ways of thinking about the same properties? It is hard to see what would qualify them as different. Thus, it turns out that adding in purposes to the characterization of perceptual and sympathetic imagining, in the manner above, does not preserve the core idea behind the distinction. Some cases of sympathetically imagining a mental state will be cognitively equivalent to perceptually imagining the same state. If there is no firewall between sympathetic and perceptual imagining—if one can in fact be an instance of the other—it is hard to see how sympathetic imagining can offer a special insight into mentality that is not offered simply by imagining the perceptible features of mental states.

In response, one might insist that we simply need to be more specific concerning the intentions with which the relevant images are used, in a way that allows the imaginings to focus on different *properties* of the (single) state being imagined. Could it not be that sympathetic imagining involves not merely the intention think about some state itself, but the more specific intention to think about the state *in terms of what it is like to have it?* (That is, in terms of its *phenomenal properties*). If so, one might then hold that, by contrast, when perceptually imagining the state, we merely intend to think about its *perceptible properties*—where perceptible properties are the sort of properties people can detect a thing as having merely

through perceiving it. The different purposes behind each kind of imagining might seem adequate to underwrite the idea that we have here two different ways imagining, with different properties as their focus—even if, in some circumstances, an imagining of each kind might incorporate the same type of image and be *of* the same state or object.

But *what it is* to aim to think about *E* in terms of what it is like to have *E* (or, equivalently, in terms of *E*'s phenomenal properties), assuming this is different than aiming to think about *E*'s perceptible properties? Here Nagel's original characterization of sympathetic imagining seemed to offer some leverage: to imagine *E* in the sympathetic way—and so to think about *E* in terms of what it is like to have it—was to think about *E* by entering into a state that resembles *E* itself, as opposed to entering into a state that resembles the state one would be in if one perceived *E*. This was an intriguing idea because it seemed to offer an independent characterization of what it is to think of a mental state in terms of what it is like to have it. It seemed to capture what Jackson's (1982) Mary cannot do with respect to experiences of red, and what we cannot do with respect to the experiences of a bat, without invoking the murky notion of *what-it's-like-ness*. But now it appears that, in order to explicate sympathetic imagining, we have to add to its characterization that it is using an image that resembles *E* itself *for the purposes of thinking about what it is like to be in E*, and not simply for the purposes of thinking about *E* (or about *E*'s perceptible properties). This means that the notion of sympathetic imagining cannot shed light on what it is to think about a mental state in terms of its phenomenal properties; instead, it relies upon an independent understanding of what it is to do so.

The force of this point can be appreciated through considering a possible objection. It may seem that differences in *how we conceptualize* phenomenal as opposed to perceptible properties could serve to distinguish the two kinds of imagining. This response accepts that thinking of a state in terms of its phenomenal properties will not be explicated by appeal to sympathetic imagination; instead, it proposes to run the explanation the other way, explaining sympathetic imagination (and phenomenal thought, generally) in terms of our ability to use images to conceptualize experiences by their phenomenal properties.

Now, if materialism is true and the explanatory gap is *merely* epistemic in nature—as Nagel, Hill, and other advocates of this approach propose—then phenomenal properties must themselves be physical properties. And if phenomenal properties are physical properties, then they can be thought about—i.e., conceptualized—through the use of neural or functional state

concepts, or perhaps even through the use of mental images of living brains (as in the case of Owen). This is why the key distinction materialists must invoke when characterizing phenomenal thought is not that between phenomenal properties and physical properties (phenomenal properties *just are* physical properties of a sort), but between thinking about phenomenal-cum-physical properties *in terms of what it is like to have them* (i.e., through use of a phenomenal concept or sympathetic imagination) and thinking about the same phenomenal-cum-physical properties in some other way (e.g., through the use of physical-state concepts or perceptual imagination). It is precisely here that the perceptual/sympathetic distinction (and the phenomenal concept/physical concept distinction) is supposed to shed light by providing a characterization of what it is to think *phenomenally* about a phenomenal-cum-physical property. (Or, equivalently, of what it is to think about a phenomenal property “as such”). There seems, to many, to be a special imagistic way of thinking about phenomenal properties, as such; appreciating this way is thought to shed light on the nature of phenomenal thought. Yet the case of Owen shows that existing accounts do not identify any special feature of such thought that is not also shared by putatively non-phenomenal thought.

Of course, the fact that these questions can be raised does not show that no elucidating account can be given for what it is to think about a state in terms of what it is like to have it. Indeed, I will offer one below. But it does clarify the task. To characterize this mode of thought—and, in the process, what it is to think about a phenomenal property “as such”—it is not enough to show how a token mental state with phenomenal character, such as a visual image, can be used in a thought about itself. For we have seen that putatively non-phenomenal forms of thought—cases where we merely think of a mental state in terms of its perceptible features—can do that as well. Instead, it must be explained how and in virtue of what it is that only *some* cases of using a mental image in a representation of itself constitute thinking about the experience in terms of what it is like to have it.

2.3 Constitutional Accounts of Phenomenal Concepts

Recent constitutional accounts of phenomenal concepts can be seen as descendants of Nagel’s notion of sympathetic imagination and might therefore be thought offer leverage on the question of why only some cases of using a mental image in a representation of itself constitute thinking about an experience in terms of what it is like to have it. These theories hold that

phenomenal concepts—i.e., those concepts that enable us to think about mental states in terms of what it is like to have them—are partly *constituted by* instances of the states or properties to which they refer (Balog, 2012; Block, 2006; Papineau, 2007b). These accounts fill in some relevant details concerning the mechanisms by which a state might come to refer to itself. However, they fail to gain ground on explaining the difference between using an image to think about itself in terms of what it is like to have it and using the same type of image to think about itself in some other way.

For instance, Katilin Balog's (2012) quotational view of phenomenal concepts echoes Nagel's conception of sympathetic imagination in holding that each phenomenal concept "refers to something that (partly) constitutes it, and refers to it *in virtue of* it being so constituted" (Balog, 2012, p. 32). Balog proposes that the device of linguistic quotation can be seen as a model for how phenomenal concepts achieve self-reference in this special way. Just as 'red' (arguably) contains an instance of the word to which it refers, so too, Balog suggests, may perceptual experiences (and mental images) come to refer to themselves through the insertion of something like quotation marks around the experience. But, whatever we make of this analogy, it serves only as an example of how a mental state might make reference to itself in virtue of being (partly) constituted by itself. We have already seen that Owen's perceptual imagining of his own brain state shares this feature: it is just because visual image E has the intentional properties it does that it represents brain state B, to which it is identical. But, presumably, Balog would not deem this an instance of phenomenal thought. Yet what is it that Owen's perceptual imagining lacks that would be *gained* if the self-reference occurred through something akin to quotation? How would gaining self-reference in that particular way *change* the properties being thought about? It is hard to see. As it stands, the appeal to quotation fares no better than Nagel's distinction in explaining how imagery can be used to think about mental properties in terms of what it is like to have them.⁹

⁹ The same points apply to Block's (2006) conception of phenomenal concepts, wherein token phenomenal properties are held to do "double duty" in both constituting and being the referent of phenomenal concepts. Block holds that phenomenal concepts have the structure: "the experience: *q*", where *q* is some phenomenal property had by the experience to which the concept refers (2006, p. 23). However, it is easy to see that when Owen sees brain state B/E during his operation, he may think: I am looking at the experience: *q*," where *q* is a phenomenal property both had by his brain state/visual experience and that he is using in the representation of his current brain state/experience. We again have a situation where the distinction collapses between using a phenomenal property to think about an experience in terms of what it's like to have it and using the same phenomenal property in an ordinary perceptual judgment about the experience.

David Papineau (2007b) offers a related but importantly different constitutional account of phenomenal concepts. Papineau holds that a particular “sensory template” can be used to represent both a worldly object (e.g. a bird) and also an experience (e.g. of seeing a bird), depending on the information it carries and the subject’s dispositions to “project” that information from encounter to encounter. (Sensory templates, for Papineau, are *roughly* equivalent to mental images.) Here Papineau explains how a certain sensory template that in one case refers to a bird can, in another, refer to an experience:

Suppose I am disposed to project, from one encounter to another, such facts as that what I am encountering ceases when I close my eyes, goes fuzzy when I am tired, will be more detailed if I go closer, and so on...then I will be referring to the experience of seeing the bird rather than the bird itself (2007b, p. 122-123).

Taking this proposal at face value, it does nothing to explain how there can be two different uses of a sensory template to imagine one and the same experience. Instead, Papineau offers an account of how, in virtue of dispositions to associate different bodies of information with different objects, a particular sensory template can have two different *referents*. In essence, his view is that one sensory template can support two distinct *mental files* (where a mental file is a repository of information one has stored about some entity). Yet, for that reason, it is not poised to offer guidance on how there can be two distinct uses of the same sensory template to imagine *the same* referent, when the person doing the imagining (such as Owen) believes he is imagining the same thing in both cases, and therefore has but one mental file for the state in question.

2.4 Looking ahead

To many it seems just obvious that there is a special imagistic form of phenomenal thought—one that is distinct from perceptual imagining and that allows us to accurately conceptualize phenomenal properties in terms of what it is like to have them (or, equivalently, to *phenomenally* conceptualize phenomenal properties). Whether there really is such a form of thought is, of course, the question at hand. More precisely, what is in question is not whether there are two distinct ways of imagistically thinking about one and the same set of conscious mental states (I grant that there are), but whether both are generally *accurate* or *veridical* ways of thinking about those mental states. If no explanatory account of what it is to use imagery to phenomenally conceptualize a phenomenal property is available to vindicate the idea that such

forms of thought are veridical, then the door is open to explanations of that form of thought that deny its veridicality.

The next section develops an approach of this kind, by rejecting the idea that there is a special imagistic way of *accurately* representing experiences in terms of what it is like to have them. To some, it will seem a disadvantage of the approach that it is a kind of error theory, according to which many imaginings misrepresent the nature of experiences. Yet my claim will be that this apparent bug is a feature, as it brings with it a diagnosis of the explanatory gap.

III. Experiences imagined, and misimagined

To make progress on the questions we have been considering, I want to propose a framework for thinking about the contents and correctness conditions of sensory imaginings in general. This is the same framework I have developed in other work (Langland-Hassan, 2015); but it will be put to much different ends here. Sensory imaginings, as I will understand them, are instances of occurrent cognition that involve mental imagery. Others have used the terms *perceptual imagining* and *sensuous imagining* to mark the same phenomenon (Byrne, 2007; Currie, 2002; Gendler, 2005; Martin, 2002; Noordhof, 2002). There are two key components to the framework. The first is that sensory imaginings have a *hybrid* content, in the sense that two kinds of representations—what we can think of as language-like and image-like representations, respectively—contribute to the content of sensory imaginings.¹⁰ Together, they ensure that the content of an imagining is a proposition with a truth value. And, second, there are different *attitudes* one can take toward such propositions—attitudes such as belief and desire—in keeping with the distinct functional roles such states may play. The only difference with propositional attitudes, traditionally conceived, is that here the content of the proposition towards which one takes an attitude is partly accounted for by the contribution of a mental image. I will use the term *imaginative attitude* to refer to the various attitudes we take toward propositions whose content is partly accounted for by the contribution of a mental image.

In expressing the contents of sensory imaginings, I will use **bold** to distinguish the specific portion of an imagining's content contributed by a sensory image (or sequence of

¹⁰ The idea that image-like and language-like (or discursive) representations combine in imagistic thought is not new. See, e.g., Fodor (1975, p. 190), Tye (1991, Ch. 5), Kung (2010), Peacocke (1985) and, indeed, both Balog (2012) and Block (2006). What I call the content of *an imagining* is what others might call the content of an *imaginative episode* or *imaginative project*, where this content involves contributions from both imagistic and non-imagistic representations.

images). Of course, the fact that psychological contents are here described in natural language should not be taken to suggest that their format is itself language-like. Rather, with respect to the contents written in **bold**, the idea is that their format is picture-like, or non-discursive. Also, it should not be assumed that, for every word included in **bold**, the related image represents the very property named. So, for instance, the image whose content is described as **a big brown horse** may not itself represent the property of being a horse. The words in **bold** are simply meant to point the reader toward a general idea of the kinds of (perhaps only low-level, superficial) properties the image represents. The properties represented by **a big brown horse** can be assumed to be whatever properties are *in fact* represented by what we would pre-theoretically describe as a mental image of a big brown horse. Likely candidates include colors, shapes, and relative spatial positions.

Finally, and importantly, I will suggest that the contents of images should be thought of as akin to *indefinite descriptions* (i.e. descriptions beginning with ‘a’ or ‘an’, or ‘some’). Among other things, this allows for an account of how one and the same image (in the sense of a type of image) can be used to imagine many different objects and scenarios.¹¹

Putting all of this into practice, let us suppose that I make an ordinary judgment that daffodils are in bloom—one that involves no mental imagery. Judgments, as I will understand them, are occurrent mental events where a person comes to have a particular belief. Using JUD to stand for the attitude of judgment, we can express the judgment that daffodils are in bloom as: JUD (daffodils are in bloom).

Similarly, we can use JIG to stand for an *imaginative* attitude corresponding to judgment—what I will call a *judgment imagining*. Judgment imaginings constitute a subset of our judgments; they are those judgments of which mental images form proper parts. We can symbolize particular judgment imaginings by using JIG with a content following in parentheses. As JIGs are a form of sensory imagination, at least some of their constituents will be sensory images, and therefore symbolized in **bold**. For instance, if a visual image that we would intuitively describe as being “of a red Prius” forms part of my judgment that Owen’s car is a red Prius, we could symbolize the sensory imagining as:

JIG (Owen’s car is **a red Prius**).

¹¹ Thus I will, *inter alia*, be offering a particular account of what is sometimes known as the Multiple Use Thesis, according to which “the same mental image may be used to fulfill different imaginative projects” (Noordhof, 2002, p. 428).

Here, ‘**a red Prius**’ is used to represent the contribution of a visual image to a judgment imagining, while ‘Owen’s car is’ represents the contribution of language-like (or discursive) representations to the imagining. The same type of image can be used in thoughts about a particular car, a type of car, a future car, a past car, a hallucination of a car, and so on, depending on variations in the non-imagistic, discursive component of the content. This is because *what is imagined*, on this account, is a separate matter from the content of the image itself; the image accounts for only part of the content of an imagining. Here are several other JIGs one might make:

- (1) JIG (Sarah wants to buy **a red Prius**).
- (2) JIG (If I owned a car that looked like Owen’s, it would be **a red Prius**).
- (3) JIG (If I were hallucinating Owen’s car, I would seem to see **a red Prius**).

Part of the idea behind imaginative attitudes is that there are many such attitudes, including conative variants, that account for the different roles imagery plays in human cognition. For instance, if Sarah is forming an image of a red Prius while fantasizing about owning a red Prius, we might characterize her occurrent desire as a desire-imagining (DIG) of the form:

- (4) DIG (Sitting in my driveway is **a red Prius**).

An important feature of this approach is that it allows us to see how we can use imagery to have thoughts that are simply *false*. For instance, suppose that, while at the furniture store, I am trying to determine whether a certain couch will fit through my doorway. I might imagine the couch and doorway in order to do so, having the thought:

JIG (As it approaches my doorway this couch would be **a tan couch fitting narrowly through a doorway**).

Now, suppose the couch arrives and does not fit through the doorway. In that case, the JIG was non-veridical. It was a judgment that turned out to be false.

We can now turn to the question of imagining experiences. Because we are interested here in the role that imagery plays in our *judgments* about experiences, I will use examples involving JIGS, as opposed to some other imaginative attitude. It will help to have a particular kind of visual image in mind. Let it again be a mental image of a beach ball, which in my notation can be symbolized as **a multi-colored smooth sphere**. Recalling Nagel’s perceptual/sympathetic distinction, it seems we can use an image of a beach ball both (perceptually) to imagine a beach ball and (sympathetically) to imagine a visual experience of a

beach ball. Put in terms of the present framework, we might perceptually imagine a beach ball with the following judgment-imagining:

(5) JIG (An ordinary beach ball is **a multi-colored smooth sphere**).

This JIG uses ‘**a multi-colored smooth sphere**’ to predicate the properties represented by the image to an ordinary beach ball. If an ordinary beach ball has those properties—a certain color, texture, and shape, say—then the JIG is veridical. Could the same type of image be used to imagine *an experience* of a beach ball? Sure. Here is one way:

(6) JIG (Yesterday I had a visual experience of **a multi-colored smooth sphere**).

This is a legitimate way to use (what we would pre-theoretically call) an image of a beach ball to think about an experience of a beach ball. A person having the thought represented in (6) could reasonably be said to be *imagining the experience* of a beach ball. After all, one is using a mental image to, *in some sense*, think about an experience. Specifically, the image in (6) helps us think about what the experience is *of* or *about*. As for a case where a mental image is used *in a thought about itself*, we can alternatively consider:

(7) JIG (I currently have a visual image of **a multi-colored smooth sphere**).

In exploiting an image in the manner of (6) and (7), we think about experiences in terms of what they are *of* or *about*. But, I take it, this is not yet to think about the experiences in terms of *what it is like* to have them. And it is not yet to think about an experience’s phenomenal properties *as such*. Why isn’t it?

For one thing, (6) and (7) are not cases of using a mental image to represent itself, or any other experience. The images in these JIGs attribute properties not to experiences, but to that which the experiences are experiences *of*. After all, (7) is more or less equivalent to:

(8) JIG (The beach ball that I have a visual image of is **a multi-colored smooth sphere**).

In (8), we are still using a mental image of a beach ball to, *in some sense*, think about the image itself. But this does not seem any closer than (5), an obvious perceptual imagining, to using an image to think about an experience in terms of what it is like to have it. What we seem to need for the latter is a case where the properties represented by the image are predicated *of the experience itself*, and not merely to whatever the experience represents. That is, we would need a JIG of the form:

(9) JIG (My current visual image is **a multi-colored smooth sphere**).

In this case, the properties represented by the image are predicated of the visual image itself. My contention is that people sometimes unwittingly make judgments of roughly this form, where properties represented by mental images are predicated of mental states themselves. However, judgments in the form of (9) are false. Humans do not have experiences that are themselves multi-colored, smooth, or spherical. We do not have experiences that have the superficial properties of beach balls. As a class, our judgement-imaginings will almost always be false when they use a mental image to predicate of an experience the properties represented by the image. If such uses are indeed what is required for us to think about our experiences in terms of what it is like to have them, then such thoughts are false.

Here one may object that that it is unfair to assume that the type of image in question—**a multi-colored smooth sphere**—only ever represents properties of the kind had by beach balls. Could it not be that, in some contexts, the same type of image represents *different* properties, of a kind had by mental states?

Well, it *could* be. This would be the case if mental images had the sort of dual roles envisioned by Nagel and defenders of constitutional accounts of phenomenal concepts. As we saw in Section 2, however, such approaches face difficult challenges in substantively characterizing the two different roles. A simpler and independently plausible approach is to hold that token mental images of any give type will always have the *same* content (and can indeed be typed by their content), involving (roughly) the kinds of perceptible properties that corresponding perceptual experiences would enable one to detect in the external environment. This approach to the contents of imagery clearly meshes with work in cognitive science that identifies neural and behavioral symmetries between visual imagery and visual perception (Ganis, Thompson, & Kosslyn, 2004; Kosslyn, Thompson, & Ganis, 2006; Slotnick, Thompson, & Kosslyn, 2005). Moreover, it is not as though this framework completely denies our ability to imagine experiences. It grants that we can imagine experiences in the sense of (7) and (8), where we imagine an experience in terms of what it is of or about. Yet the framework does entail that imaginings in the form of (9) will—*almost* always—be non-veridical. That such imaginings are ruled non-veridical is an advantage of the account, however, as it allows for an explanation of how there comes to be the appearance of an explanatory gap. Precisely *how* it allows for that explanation is the topic of Section 4.

Let us return now to the question that faced both Nagel's sympathetic/perceptual imagination distinction and constitutional accounts of phenomenal concepts: what is the difference between using an image to think about an experience in terms of what it is like to have it, and using the same type of image to think about the experience's perceptible features? Owen, in our earlier example, forms a visual image of the neural regions that realize the very visual image he is forming. This visual image represents the perceptible features of a particular brain state, such as its color, shape, and texture. To symbolize the image and its content, we can use: **a pink and red bit of neural tissue**.

The interesting feature of Owen's Operation is that it is the kind of exceptional case where a person can *correctly* use a mental image to predicate of the mental image itself the very properties it represents, even on the supposition that mental images only represent perceptible properties. An example of such a thought, as had by Owen, would be:

(10) JIG (My current visual image is **a pink and red bit of neural tissue**).

Let us suppose that Owen's judgment made via (10) is true. Is (10) an instance of thinking of a visual image in terms of what it is like to have it? It does, after all, involve the use of an image to predicate properties represented by the image to itself. Or is it merely a case of thinking of a visual image in terms of its perceptible features? (One thing is clear: it is not, like (7) or (8) a case of thinking of one's image in terms of what it is *of* or *about*.)

I think the correct response here is that (10) could be the thought one is having *either* when one thinks about a certain visual image in terms of what it is like to have it, *or* when one thinks of it in terms of its perceptible features. What makes it the case that a person is thinking about the experience in one or the other way lies not in the content or the attitude of the judgment itself—these can be the *same* in both cases—but in the inferential pattern out of which it arises. By an "inferential pattern" I simply mean a habit of thought, a tendency to use mental imagery in one of two ways. There is a pattern—call it the *what it's like* pattern—that consists in judgments such as (9), where the properties represented by mental images are predicated of experiences themselves. Given that images themselves just represent the kinds of properties we can perceive, all instances of this pattern that do not involve images of perceptible features of brains will be false (for us, at least); human experiences will simply lack the properties attributed by the image. Yet there is also another pattern—call it the *perceptible features* pattern—exemplified by cases such as (5), where the properties represented by a

mental image are predicated of some object or entity of the sort that we are able to perceive, and to perceive as having the kind of properties attributed by the image.¹²

The what it's like pattern of inference is not a *safe* pattern of inference, insofar as most judgments of its type are false. Thus, even if the JIG represented in (10) constitutes a true judgment, we can reasonably hold that it does not amount to knowledge if it arises out of the epistemically unsafe *what it's like* pattern. On the other hand, if (10) arises out of the *perceptible features* pattern of inference, it may well qualify as an instance of knowledge. For we are more or less reliable in our uses of mental imagery to represent the kinds of things we can ordinarily perceive (Finke, 1989; Kosslyn et al., 2006).

IV. Lights, Thought Bubbles, and the Explanatory Gap

Up to this point, I have developed a challenge to the idea that there is a special (and veridical) imagistic way of thinking about experiences (Section 2) and argued for a simpler alternative conception of what is involved in imagining experiences (Section 3). According to this alternative, the sort of use to which many are inclined to put imagery when thinking of experiences in terms of what it is like to have them is a *misuse*. It involves attributing to experiences a range of perceptible properties they almost invariably lack. In this section I want to defend the idea that this kind of misuse contributes to the appearance of an *explanatory gap*.

The most general reason that imaginings arising out of the what-it's-like pattern create the appearance of an explanatory gap is that they involve a person's (unwittingly) attributing to experiences properties unlike any they actually possess. If one is indeed committed, on the basis of such imaginings, to experiences having those properties, it will be hard to see how we might capture them in a reductive account of mind. Suppose, for instance, that one were to use the image **a technicolor raincoat** in order to think about the visual experience of seeing a technicolor raincoat (in terms of what it is like to have the experience). One would be attributing to a certain visual perceptual state a rich array of colors, shapes, and textures that it lacks. The idea that a mental state, so characterized, might be identical to a certain kind of neural activity would be very puzzling indeed.

¹² There is also a third pattern, exemplified by (7) and (8), which we might call the *intentional object pattern*, where we use images to represent the perceptible features of the intentional object of some mental state.

Now, it is important to note that one might not *explicitly realize* that these were the properties being attributed by one's images. Imagistic representations are non-conceptual and non-linguistic in nature, which leaves open the possibility that a person may mischaracterize, at the level of verbal report and conceptual understanding, the nature of the properties they are thinking about through use of the image. All we should expect from those making such a mistake is that they should judge their experience to have *colorish*, *shapeish*, and *textureish* properties. (They might do so even if they agreed it did not make much sense to say that their experiences were *literally* colored, shaped, and textured in those ways.)

And, indeed, such a tendency is evident in the specific metaphors used to invoke the explanatory gap. Consciousness is described as a "mysterious flame" (McGinn, 2000). In its absence, "all is dark inside" (Chalmers, 1996, p. 96). A central task for philosophical theories of consciousness is to explain how conscious mental states "light up" (Rosenthal, 2004, p. 19). Esteemed neuroscientists describe the acquisition of consciousness as "stepping into the light," (Damasio, 1999) and agree that information-processing in the absence of consciousness must occur wholly "in the dark" (Koch, 2004, p. 231).

Surely it is reasonable to ask *why* light metaphors in particular seem appropriate to the characterization of consciousness and its absence. Why not wind or heat? Why is consciousness not well described as a "mysterious breeze"? Why is there no need for a theory of what makes conscious mental states "warm up"? Answer: like visual perceptual experiences, visual images represent well-lit environments. They do not, after all, represent unilluminated, non-visible colors and objects. Visual images represent colors and objects that are more or less suffused in light. Thus, when we imagine experiences in line with the what-it's-like pattern, we inadvertently imagine a scenario suffused in light. We are left with the strong sense that there is something luminescent or well-lit about experiences, so imagined. Like "tropical sea-water," they appear "phosphorescent" and "self-luminous" (Ryle, 1949, p. 159). Of course no one believes that experiences are *literally* illuminated or aflame; yet still the metaphors seem somehow *appropriate*. The present error-theory of imagining experience explains this particular *character* of the explanatory gap, insofar as it predicts the kind of metaphors used to invoke it.

It would nevertheless be of dialectical aid if some account could be given of *why* people would be inclined to misrepresent their experiences in this way in the first place, and why—even after having the putative error pointed out—one might still feel convinced that there is

nevertheless something truth-preserving about imagining experiences in this way. To that end, I want to explore a possibility that links the attractiveness and usefulness of misrepresenting experiences in this way to our means for monitoring other minds and predicting others' behavior. This requires a brief digression on the topic of thought bubbles.

4.1 Thought bubbles and theory of mind

Thought bubbles are little clouds above the heads of depicted thinkers, wherein the contents of their thoughts are shown. Thought bubbles are helpful when an artist wants to convey that the character is thinking about something not otherwise depicted, or to show that the thoughts of one character differ from those of another. To make use of a thought bubble, the artist draws the very thing being thought about and then draws a bubble around it. Note that the line constituting the circumference of the bubble does not *depict* anything. It is instead a symbolic device that, by its location on the page (e.g., near Garfield's head), tells us *who* is thinking about the thing inside the bubble (e.g., lasagna) (see Fig. A).

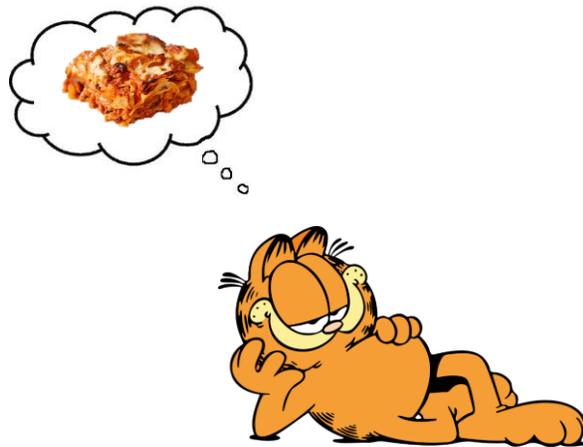


Figure A – Garfield's wish

What about the picture inside the bubble? What does the picture *depict*? A piece of lasagna? An experience? Depending on how we interpret the semantics of the bubble itself, and drawing on our earlier framework (from Section 3), there are a number of possibilities for what the cartoon as a whole should be understood as expressing, including:

(11) Garfield is thinking about **a piece of lasagna**.

Or:

(12) Garfield's thought is **a piece of lasagna**.

Or, combining the two:

(13) Garfield is thinking about **a piece of lasagna** by means of **a piece of lasagna**.

According to (11), the contents of the bubble depict what the character is thinking of.

According to (12), the contents of the bubble depict the character's thought itself. And (13) generously views both (11) and (12) as possessing a grain of truth.

I recommend (11) as the most plausible semantics for thought bubbles. Trading as they do on the what-it's-like pattern of inference, (12) and (13) involve explicit falsehoods. But why is it nevertheless tempting to think that the image in the bubble depicts Garfield's thought itself, and not merely a piece of lasagna? The reason, I want to suggest, is the *same reason* that the what-it's-like pattern of imagining experiences has intuitive appeal. The general tendency to think that imagery can be used to depict minds derives from the fact that we use a cognitive equivalent to thought bubbles in keeping track of other minds—in representing, for instance, how another's point of view differs from our own. If, for example, Susan has a view of something I cannot see—she's peering around a corner to look at fancy car, say—I might form a visual image of what I think she is seeing and associate the image with Susan's head. Associating the mental image of a sports car with Susan's head is an easy way to keep track of her cognitive situation when it differs from mine. Or, when playing poker, I might associate a different image with each of the other players—an image that captures what I suspect each can see from his or her position. Because each player will have a different associated image, and because the point of each image is to record what each player can see, it will be natural to associate the image with what is going on in the other player's mind. I might even imaginatively *project* each image into the scene I am perceiving, “placing” each image by the associated person's head, just as we more generally project imagery onto perceived scenes in order to plan our actions (Briscoe, 2008; Van Leeuwen, 2011). And, in one sense, this image-projection will be correct: each image represents what I think the associated player's visual experience is *of or about*. But, interpreted slightly differently, it is incorrect: the properties represented by the image are not properties had by anything inside the players' heads.

This is not to say that the process of associating images with other people's minds is something that we consciously and methodically do in every case; nor is it our only means for mental bookkeeping. My claim is that it is one of many mindreading heuristics we come to employ in an automatic and reflexive manner. It is precisely the involuntary and unconscious manner of its application, and the fact that it forms part of an otherwise reliable practice of

projecting imagery onto perceived environments, that accounts for the deep-seated conviction that there is something obviously *right* about using visual and other sensory images to represent mental states themselves.

This philosophical hunch has ample support from developmental psychology. Wellman *et al.* (1996) found that, after a brief training period, even three year old children (who do not reliably pass traditional false-belief tasks (H.M. Wellman, Cross, & Watson, 2001)) were able to understand thought bubble pictures as showing the thoughts of different individuals (see also Kerr & Durkin (2004)). This included their understanding a situation where two characters had *different* thoughts about the identity of a single object inside an opaque box. Two things are striking about the study: first, it highlights the *conventional* nature of thought bubbles, insofar as almost all of the three and four year olds had to be taught to interpret the bubbles as showing what the characters are thinking. This serves as a useful reminder that thought bubbles are no less arbitrarily related to their meanings than the dashes and “action lines” used in cartoons to convey movement or sound. Second, this means for representing thoughts was both *easy* for young children to learn and *powerful*, enabling them to quickly show an understanding of how the thoughts of one person can conflict with those of another.

Additional evidence of the power and ease of this means for thinking about other minds comes from work on individuals with theory of mind deficits. Children with autism spectrum disorder (ASD) are well-known for having impaired social cognition, passing false belief tasks considerably later than typically developing children of comparable verbal age (Baron-Cohen, 1989). Yet multiple studies have shown that if children with ASD are trained in the conventions of using thought bubbles or “pictures in the head” to represent mental states, they significantly improve at passing false-belief tasks, including versions of those tasks they have not previously encountered (see, e.g., Wellman *et al.* (2002), Swettenham *et al.*, (1996), McGregor *et al.* (1998), Paynter & Peterson (2013), and Parsons & Mitchell (1999)). In these studies, participants first learn to pass a false belief task when thought bubbles showing the contents of the characters’ thoughts are present above the heads of each character, and are subsequently able to pass new versions of the tests without the thought bubbles present. Paynter & Peterson (2013) recently replicated Wellman *et al.*’s results, showing that performance remained improved even three weeks after training.

The ease of acquisition and power of thought-bubble-thinking to both typically developing children and children with ASD make it plausible that ordinary adults adopt a similar

heuristic *very early* in the development of their own theories of mind. And the nature of the heuristic—projected images being closely *spatially* associated with minds—offers a way of seeing why, when it becomes part of our implicit and reflexive theory of mind, we feel a strong pull to the idea that mental images can be used to depict experiences themselves. After all, most of us have usefully done so our entire lives. There subsequently emerges an ambiguity, reflected in JIGs (11)-(13), concerning just what the pictures inside thought bubbles represent, and how. The close association of an image with a person's mind makes it tempting to assume that the image represents the person's thoughts in the same way pictures represent other things—by *depicting* them. At the same time, an image of lasagna will appear to be just that—*of lasagna*. We might then straddle the fence by loosely conceiving of the thought bubble as representing that a character is thinking about that which is depicted in the bubble, *by means of something* that is *itself* accurately depicted by the image in the bubble (as in (13) above). Yet the second half of this assumption will give rise to deep perplexities concerning the mind-brain relation, for those inclined to reflect. How can something correctly depicted by an image of a beach ball be identical to something that is also correctly depicted by an image of a brain state? That was Nagel's question, to which a revolutionary answer seemed necessary.

A better alternative is to resist thinking of images as depicting experiences in the first place. The usefulness of the mental bookkeeping heuristic does not require it. The heuristic only requires that we associate an image with a person's state of mind, where the image depicts what the person is thinking *about*. It does not require that we take the image to depict the person's mental state itself. The latter assumption can be discarded while the heuristic retains its use. At the same time, the fact that we have, since early childhood, wrongly conceived of images as depicting experiences themselves serves to explain the deep-seated intuition that images can, in fact, accurately depict experiences.

To some, this account of the mistake made in using imagery to represent mental states will seem to get matters back to front. One might object that the reason we find it plausible (if we do) that images depict experiences themselves derives from introspection—that is, from direct awareness of *our own* minds—and not from a heuristic for understanding *other* minds. While there is no denying the pull of such an intuition, it bears noting how little consensus there is concerning the kinds of properties we attend to when introspectively attending to our own experiences. As against the intuition just mentioned, a contrary intuition is that our

perceptual experiences are *transparent* in the sense that, when we try to introspectively attend to them, we only attend more closely to the external objects and properties they represent (Byrne, 2012; Dretske, 2003; Harman, 1990; Tye, 2009). If those who favor transparency are correct concerning the deliverances of introspection, then the intuition that images can be used to depict or otherwise represent experiences must have some source *other than* introspection. Appealing to the above theory of mind heuristic to explain the (mistaken) intuition then becomes an attractive option, not least because it forms part of a larger explanation for the appearance of an explanatory gap.

One might also worry that the present proposal attributes a rather remarkable and profound error to the way that many of us think about minds—too profound, one might think, to be plausible. Yet the level of mistake is no greater than that which must attend to one side or other of the transparency debate. One side holds that when we introspectively attend to our experiences we are aware of features of our own minds; the other maintains that we are aware of external objects and their properties. That is a considerable disagreement. Whichever side we choose, we cannot avoid holding that many good minds are quite seriously mistaken—mistaken about something that, seemingly, *ought* to be obvious. Thus it is no mark against the present argument that it attributes to many a deep error. The important point is that there is an explanation for the error, and appreciating the error itself helps to narrow the explanatory gap.¹³

5. Conclusion

A particular kind of misuse of imagery is to blame for at least some of the sense of contingency surrounding the mind-brain relation. It is a misuse that consists in using mental images to (unwittingly) predicate the properties represented by images to experiences themselves—or so I have argued. Adopting this perspective allows for an explanation of the difference between using an image to think about an experience in terms of what it's like to have it, and using the same type of image to think about the same experience in terms of its

¹³ As noted at the outset, I do not think there is just one reason people are inclined to intuit an explanatory gap. The perplexities arising out of the literature on personal identity, for instance, may have a slightly different root, and may also lead one to intuit a gap (Parfit, 1984). Thus we should not be surprised if, even accepting the error theory of imagination advocated here, some still sense that there remains an explanatory gap that has not been addressed.

perceptible features. This is a distinction that has not been explained by views that assign a more substantive and veridical role to imagery in thought about experiences. The failure of such views to capture this distinction betrays a more general lack of clarity concerning the nature of this putatively special form of phenomenal thought. With the positive account of the distinction proposed here, we arrive at a clear explanation of what it is, *in general*, to use imagery to think about experience in terms of what it is like to have them. Further, it is from within the terms of this error-theory of imagining experiences that we arrive at a plausible explanation for the appearance of an explanatory gap. In so imagining experiences, we are attributing to experiences properties unlike any they actually possess. Finally, this account of the appearance of the explanatory gap has independent support when we consider the role of imagery in our (likely implicit) theory of mind heuristics. The mistake in question could easily result from the reflexive use of certain powerful and easy-to-learn strategies for mental bookkeeping.

In exploiting this strategy to close the mind-brain gap, however, one might worry that another gap opens: the color-body gap. As Shoemaker (1996, pp. 248-249), Byrne (2006), and others have remarked, it is at least *possible* to intuit a gap between colors *as we visually perceive them*, and colors *as described by color science* (assuming here that colors are physical properties of ordinary objects). Thus, even if one were to accept the present dissolution of the mind-body explanatory gap, the color-body problem arguably remains (Byrne, 2006). After all, before leaving her black and white room, Jackson's Mary can no more imagine *red itself* than she can imagine the experience of red. Whatever we make of *that* inability, the questions it raises cannot explained away by any means proposed here.

A few brief points in response: first, the color-body problem—if it is a problem—is more obviously a problem about *how we think about* certain properties than a problem about the *special kinds of properties we think about* when we think about colors (or minds). (Here I agree with Byrne (2006, p. 243)). Setting aside all talk of *what-it's-like-ness*, the key question, very roughly, is whether two *a priori* inequivalent ways of thinking about something entail two different properties of the thing being thought about. Some discussions of the mind-body problem identify this as the key question there as well; and, indeed, many of the points made by Block (2006) in the context of fending off mind-body property dualism can be transferred directly to the color-body problem.

Second, delicate issues concerning co-referential concepts aside, we take a step forward simply in moving the question outside of the mind. For the mind body problem is not just *one* problem, but a hornet's nest of overlapping yet distinct questions concerning intentionality, personal identity, free will, knowledge, and consciousness. To the extent that all of these questions appear to concern the same thing—the mind and its properties—they serve to reinforce each other. Where some can be reframed as questions concerning non-mental phenomena, we should welcome the opportunity.

Finally, and most importantly, if there really is a quite general explanatory gap between the external world *as perceived*, on the one hand, and *as investigated by the sciences*, on the other, it is doubtful that such a gap deserves the name. Where gaps are the rule, it is their exceptions that call for explanation.

References

- Armstrong, D. (1962). *Bodily Sensations*. London: Routledge and Kegan Paul.
- Bain, D. (2007). The location of pains. *Philosophical Papers*, 36(2), 171-205.
- Balog, K. (2012). Acquaintance and the mind-body problem. In S. Gozzano & C. S. Hill (Eds.), *New Perspectives on Type Identity: The Mental and the Physical* (pp. 16-42): Cambridge University Press.
- Baron-Cohen, S. (1989). The autistic child's theory of mind: a case of specific developmental delay. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 30(2), 285-297.
- Block, N. (2006). Max Black's objection to mind-body identity. *Oxford Review of Metaphysics*, 3, 3-78.
- Briscoe, R. (2008). Vision, Action, and Make-Perceive. *Mind and Language*, 23(4), 457-497.
- Byrne, A. (2006). Color and the Mind-Body Problem. *Dialectica*, 60, 223-244.
- Byrne, A. (2007). Possibility and Imagination. *Philosophical Perspectives*, 21(1), 125-144.
- Byrne, A. (2012). Knowing What I See. In D. Smithies & D. Stoljar (Eds.), *Introspection and Consciousness* (pp. 183-210). Oxford: Oxford University Press.
- Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.
- Currie, G. (2002). Imagination, Delusion and Hallucination. *Mind and Language*, 15(1), 168-183.
- Damasio, A. (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Orlando, FL: Harcourt.
- Doggett, T., & Stoljar, D. (2010). Does Nagel's footnote eleven solve the mind-body problem? *Philosophical Issues*, 20(1), 125-143.
- Dretske, F. (2003). How do you know you are not a zombie? In B. Gertler (Ed.), *Privileged Access: Philosophical Accounts of Self-Knowledge* (pp. 1--14): Ashgate.
- Finke, R. A. (1989). *Principles of Mental Imagery*. Cambridge, MA: MIT Press.
- Fodor, J. A. (1975). *The Language of Thought*. New York: Crowell.
- Ganis, G., Thompson, W. L., & Kosslyn, S. (2004). Brain areas underlying visual mental imagery and visual perception: an fMRI study. *Cognitive Brain Research*, 20, 226-241.
- Gendler, T. (2005). Imagination (Addendum). In D. Borchert (Ed.), *Encyclopedia of Philosophy* (2 ed., Vol. 4, pp. 599-602). Detroit: Macmillan Reference USA.
- Harman, G. (1990). The intrinsic quality of experience. *Philosophical Perspectives*, 4(n/a), 31-52.
- Hill, C. (1997). Imaginability, conceivability, possibility and the mind-body problem. *Philosophical Studies*, 87, 61-85.
- Hill, C., & McLaughlin, B. P. (1999). There are fewer things in reality than are dreamt of in Chalmers's Philosophy. *Philosophy and Phenomenological Research*, 59(2), 445-454.
- Jackson, F. (1982). Epiphenomenal Qualia. *The Philosophical Quarterly*, 32(127), 127-136.
- Kerr, S., & Durkin, K. (2004). Understanding of thought bubbles as mental representations in children with autism: implications for theory of mind. *J Autism Dev Disord*, 34(6), 637-648.
- Koch, C. (2004). *The Quest for Consciousness*. Englewood, CO: Roberts & Company.
- Kosslyn, S., Thompson, W. L., & Ganis, G. (2006). *The Case of Mental Imagery*. New York: Oxford University Press.
- Kung, P. (2010). Imagining as a guide to possibility. *Philosophy and Phenomenological Research* 81(3), 620-633.
- Langland-Hassan, P. (2015) Imaginative attitudes. *Philosophy and Phenomenological Research* 90 (3), 664-687.
- Levin, J. (2007a). Nagel vs. Nagel on the Nature of Phenomenal Concepts. *Ratio*, 20(3), 293-307.

- Levin, J. (2007b). What is a phenomenal concept? In T. Alter & S. Walter (Eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism* (pp. 87-110): Oxford University Press.
- Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, 64(October), 354-361.
- Loar, B. (1997). Phenomenal states (second version). In (N. Block, O. Flanagan, & G. Güzeldere, eds). In O. J. Flanagan, N. Block, & G. Guzelde (Eds.), *The Nature of Consciousness*: Mit Press.
- Martin, M. G. F. (2002). The Transparency of Experience. *Mind and Language*, 17(4), 376-425.
- McGinn, C. (2000). *The Mysterious Flame: Conscious Minds in a Material World*. New York, NY: Basic books.
- McGregor, E., Whiten, A., & Blackburn, P. (1998). Transfer of the Picture-in-the-Head Analogy to Natural Contexts to Aid False Belief Understanding in Autism. *Autism*, 2(4), 367-387. doi:10.1177/1362361398024004
- Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, 83(4), 435-450.
- Noordhof, P. (2002). Imagining objects and imagining experiences. *Mind and Language*, 17(4), 426-455.
- O'Sullivan, B., & Schroer, R. (2012). Painful Reasons: Representationalism as a Theory of Pain. *Philosophical Quarterly*, 62(249), 737-758.
- Papineau, D. (2002). *Thinking About Consciousness*. Oxford: Oxford University Press.
- Papineau, D. (2007a). Kripke's Proof is Ad Hominem Not Two-Dimensional. *Philosophical Perspectives*, 21, 475-494.
- Papineau, D. (2007b). Phenomenal and Perceptual Concepts. In T. Alter & S. Walter (Eds.), *Phenomenal Concepts and Phenomenal Knowledge*. Oxford: Oxford University Press.
- Parfit, D. (1984). *Reasons and persons*. Oxford Oxfordshire: Clarendon Press.
- Parsons, S., & Mitchell, P. (1999). What Children with Autism Understand about Thoughts and Thought Bubbles. *Autism*, 3(1), 17-38. doi:10.1177/1362361399003001003
- Paul, L. A. (2014). *Transformative Experience*. Oxford: Oxford University Press.
- Paynter, J., & Peterson, C. C. (2013). Further evidence of benefits of thought-bubble training for theory of mind development in children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, 7(2), 344-348. doi:<http://dx.doi.org/10.1016/j.rasd.2012.10.001>
- Peacocke, C. (1985). Imagination, Possibility and Experience. In J. Foster & H. Robinson (Eds.), *Essays on Berkeley: a Teracentennial Celebration*. Oxford: Clarendon Press.
- Rosenthal, D. M. (2004). Varieties of higher-order theory. In R. Gennaro (Ed.), *Higher-order theories of consciousness: an anthology* (pp. 17-44). Amsterdam: John Benjamins.
- Ryle, G. (1949). *The Concept of Mind*. New York: Barnes & Noble.
- Shoemaker, S. (1996). *The First-Person Perspective and Other Essays* (Vol. 108): Cambridge University Press.
- Slotnick, S., Thompson, W., & Kosslyn, S. (2005). Visual mental imagery induces retinotopically organized activation of early visual areas. *Cerebral Cortex*, 15, 1570-1583.
- Stoljar, D. (2006). *Ignorance and Imagination*. Oxford: Oxford University Press.
- Swettenham, J. G., Baron-Cohen, S., Gomez, J. C., & Walsh, S. (1996). What's inside someone's head? Conceiving of the mind as a camera helps children with autism acquire an alternative to a theory of mind. *Cogn Neuropsychiatry*, 1(1), 73-88. doi:10.1080/135468096396712
- Tye, M. (1991). *The imagery debate*. Cambridge, Mass.: MIT Press.
- Tye, M. (2005). Another look at representationalism and pain. In M. Aydede (Ed.), *Pain: New Essays on its Nature and the Methodology of its Study*: Mit Press.

- Tye, M. (2009). *Consciousness Revisited: Materialism Without Phenomenal Concepts*: Mit Press.
- Van Leeuwen, N. (2011). Imagination is Where the Action Is. *Journal of Philosophy*, 108(2), 55-77.
- Wellman, H. M., Baron-Cohen, S., Caswell, R., Gomez, J. C., Swettenham, J., Toye, E., & Lagattuta, K. (2002). Thought-bubbles help children with autism acquire an alternative to a theory of mind. *Autism*, 6(4), 343-363.
- Wellman, H. M., Cross, D., & Watson, J. K. (2001). A meta-analysis of theory of mind: The truth about false belief. *Child Dev*, 72, 655-684.
- Wellman, H. M., Hollander, M., & Schult, C. A. (1996). Young children's understanding of thought bubbles and of thoughts. *Child Dev*, 67(3), 768-788.
- Yablo, S. (2006). No fool's cold: notes on illusions of possibility. In J. Macià & M. García-Carpintero (Eds.), *Two-Dimensional Semantics* (pp. 327-380). Oxford: Oxford University Press.